Special on Survey Methodology: What is the price of going fully online? A mixed-mode sampling experiment on the occasion of the Swiss general elections 2019

Milic, Thomas ; Serdült, Uwe ; Brüggemann, Salim

Abstract: The FOKUS Aargau Mixed Mode Survey Experiment Survey-based research is currently in the process of moving partly or completely online. The scalability of online surveys not only helps to cut costs but also allows for a streamlined production of reporting which draws directly on the data. Data pre-processing can be cut to a minimum or be largely eliminated. Whereas current experience with online surveys in Switzerland shows good response rates, less is known about the quality of the sample in comparison to telephone or postal surveys. On the occasion of the October 2019 general elections in Switzerland, the FOKUS Aargau team at the ZDA therefore conducted an experiment evaluating how the response rate as well as the sample quality might differ in an online versus an offline survey. Based on the official address register, 10'000 randomly drawn residents of the Canton of Aargau who are also Swiss citizens received an invitation letter asking them to take part in the survey. While half of the sample received a paper version as well as a pre-paid envelope, giving them the option to submit their questionnaires by mail, the other half was directed to the address of the online survey with a provided link (or QR-code) only. Results The initial as well as total response rate among survey respondents who had the choice of either using the paper version and sending it in via pre-paid mail or using the online version was somewhat higher than for the sample in which the only available channel with which to respond was the internet. However, the difference in response rates between the two modes is not large and one might conclude that the paper mode is more or less obsolete. However, response rates are not the only criterion when evaluating the quality of a survey sample. With regard to representativeness the two sub-samples tend to complement each other. The sample with the mail-in option is more representative of gender, older age groups and people with a lower formal level of education. For the online sample the pattern is the reverse. In total, the mixed-mode samples, in combination, clearly yielded the most representative sample. Regarding over-reporting, there is no difference between the two survey modes. The most striking differences between respondents' answering mode are revealed when looking at the political preferences, which are crucial for a post-election survey. Mail-in survey respondents clearly tend to favor identifying and voting for the SVP (Swiss People's Party), online mode respondents more often vote for left-wing political parties. However, when controlling for other variables in a multi-variate model, party-based differences almost disappear and what emerges in its place is a pattern of digital divide, e.g. with respondents favoring the nationwide introduction of internet voting tending to opt for the online survey mode. In sum, consistent with other studies, we find that the exclusion of offline households produces significant coverage biases, while the inclusion of these households in the sample improves the representativeness of the survey significantly. Non-response as the last of the three quality criteria compared among the two survey modes showed stark differences. With a meagre completion rate of 52 percent, the paper version fares much worse than the online mode for this survey (84% completion).

Zentrum für Demokratie Aarau

Centre for Research on Direct Democracy c2d | UZH
Allgemeine Demokratieforschung | UZH
Politische Bildung und Geschichtsdidaktik | PH FHNW

Thomas Milic, Uwe Serdült, Salim Brüggemann

FOKUS Aargau

Special on Survey Methodology:

What is the price of going fully online?

A mixed-mode sampling experiment on the occasion of the Swiss general elections 2019

February 2019

specials.fokus.ag
www.zdaarau.ch

SWISSLOS Kanton Aargau

publitest
publitest

Contents

Dieses Dokument wurde zuletzt geändert am 13. November 2020. Die aktuellste Fassung ist unter folgendem Link erhältlich: fokus.ag/pdf_special_2

Zusammenfassung

Das FOKUS-Aargau-Mixed-Mode-Umfrageexperiment

Die Umfrageforschung bewegt sich heutzutage immer stärker in Richtung online. Das liegt nicht nur an den tieferen Kosten von web-basierten Umfragen, sondern auch an der effizienteren Aufbereitung von Daten und Berichterstattung. Daten und Elemente der Berichterstattung wie Tabellen und Grafiken sind dank digitaler Verarbeitung direkt integrier- und wiederverwendbar. Erfahrungen in der Schweiz mit Online-Umfragen zeigen gute Rücklaufquoten. Über die Qualität von Stichproben bei Online-Befragungen im Vergleich zu Lösungen auf Papier mit Postversand ist jedoch relativ wenig bekannt. Anlässlich der Parlamentswahlen vom Oktober 2019 in der Schweiz hat das Team von FOKUS Aargau am ZDA deshalb ein Experiment durchgeführt, in dem untersucht wurde, wie sich die Rücklaufquote sowie die Qualität der Stichprobe bei einer Online- gegenüber einer Offline-Umfrage unterscheidet. Basierend auf dem offiziellen Adressregister erhielten 10'000 zufällig ausgewählte Schweizer Bürgerinnen und Bürger des Kantons Aargau ein Einladungsschreiben mit der Bitte, an der Umfrage teilzunehmen. Während die Hälfte der Stichprobe eine Papierversion erhielt und die Möglichkeit hatte, ihren jeweiligen Fragebogen mittels eines frankierten Umschlags per Post einzureichen, wurde die andere Hälfte nur mit einem Link (oder QR-Code) auf die Adresse der Online-Umfrage hingewiesen.

Resultate

In dieser Studie betrachten wir drei Kriterien, worin sich online von auf Papier ausgefüllten Fragebogen unterscheiden könnten:

1. Der Unterschied in der **Rücklaufquote** zwischen den beiden Modi ist nicht sehr stark ausgeprägt (mit nur leichten Vorteilen für die gedruckte Variante), und man könnte versucht sein, den Schluss zu ziehen, dass der Papiermodus mehr oder weniger veraltet ist. Die Rücklaufquote ist jedoch nicht das einzige Kriterium für die Bewertung der Qualität einer Stichprobe.

2. Hinsichtlich **Repräsentativität** ergänzen sich die beiden Unterstichproben in der Tendenz deutlich. Die Stichprobe mit der Mail-In-Option ist repräsentativer für das Geschlecht, ältere Altersgruppen und Personen mit einem niedrigeren formalen Bildungsniveau. Bei der Online-Stichprobe ist das Muster umgekehrt. Insgesamt ergab die Kombination beider Modi eindeutig die repräsentativste Stichprobe. Die auffälligsten Unterschiede zwischen den Antwortmodi der Befragten zeigen sich vorderhand, wenn man die politischen Präferenzen betrachtet. Bei einer Mail-in-Umfrage tendieren die Befragten stärker dazu, sich mit der SVP zu identifizieren, während die Befragten im Online-Modus häufiger für linke politische Parteien stimmen. Sobald in einem multi-variaten Modell zur Kontrolle weitere Variablen hinzugezogen werden, verschwinden die parteibasierten Unterschiede jedoch nahezu und zeigen ein Muster der digitalen Spaltung auf, z. B. bevorzugen Befragte, die eine landesweite Generalisierung von E-Voting befürworten, eher den Online-Umfragemodus. Zusammenfassend und in Übereinstimmung mit anderen Studien finden wir, dass der

Ausschluss von Antwortenden, welche den Print-Fragebogen bevorzugen, signifikante Verzerrungen erzeugen würde, während die Einbeziehung dieser Personen in die Stichprobe die Repräsentativität der Umfrage deutlich verbessert.

3. Die **Nichtbeantwortung von Fragen** als letztes der drei Qualitätskriterien im Vergleich zwischen den beiden Erhebungsmodi zeigte starke Unterschiede. Mit einer mageren Ausfüllungsrate von 52 Prozent schneidet die Papierversion bei dieser Umfrage deutlich schlechter ab als der Online-Modus (84 % komplett ausgefüllte Fragebogen). Die automatische Führung durch den Fragebogen sowie die Hinweise, im Fall eine Frage nicht beantwortet wurde, resultieren bei der Online-Version offensichtlich in deutlich vollständiger ausgefüllten Interviews.

Abstract

The FOKUS Aargau Mixed Mode Survey Experiment

Survey-based research is currently in the process of moving partly or completely online. The scalability of online surveys not only helps to cut costs but also allows for a streamlined production of reporting which draws directly on the data. Data pre-processing can be cut to a minimum or be largely eliminated. Whereas current experience with online surveys in Switzerland shows good response rates, less is known about the quality of the sample in comparison to telephone or postal surveys. On the occasion of the October 2019 general elections in Switzerland, the FOKUS Aargau team at the ZDA therefore conducted an experiment evaluating how the response rate as well as the sample quality might differ in an online versus an offline survey. Based on the official address register, 10'000 randomly drawn residents of the Canton of Aargau who are also Swiss citizens received an invitation letter asking them to take part in the survey. While half of the sample received a paper version as well as a pre-paid envelope, giving them the option to submit their questionnaires by mail, the other half was directed to the address of the online survey with a provided link (or QR-code) only.

Results

The initial as well as total response rate among survey respondents who had the choice of either using the paper version and sending it in via pre-paid mail or using the online version was somewhat higher than for the sample in which the only available channel with which to respond was the internet. However, the difference in response rates between the two modes is not large and one might conclude that the paper mode is more or less obsolete. However, response rates are not the only criterion when evaluating the quality of a survey sample. With regard to representativeness the two sub-samples tend to complement each other. The sample with the mail-in option is more representative of gender, older age groups and people with a lower formal level of education. For the online sample the pattern is the reverse. In total, the mixed-mode samples, in combination, clearly yielded the most representative sample. Regarding over-reporting, there is no difference between the two survey modes. The most striking differences between respondents' answering mode are revealed when looking at the political preferences, which are crucial for a post-election survey. Mail-in survey respondents clearly tend to favor identifying and voting for the SVP (Swiss People's Party), online mode respondents more often vote for left-wing political parties. However, when controlling for other variables in a multi-variate model, party-based differences almost disappear and what emerges in its place is a pattern of digital divide, e.g. with respondents favoring the nationwide introduction of internet voting tending to opt for the online survey mode. In sum, consistent with other studies, we find that the exclusion of offline households produces significant coverage biases, while the inclusion of these households in the sample improves the representativeness of the survey significantly. Non-response as the last of the three quality criteria compared among the two survey modes showed stark differences. With a meagre completion rate of 52 percent, the paper version fares much worse than the online mode for this

survey (84 % completion).

1 Introduction

As survey response rates in general are in decline and, perhaps more importantly, costs for telephone surveys are comparatively high (e.g. Blumberg and Luke 2007; Keeter et al. 2006; Kohut et al. 2012), opinion researchers in the public and private sector alike are increasingly turning to web-administered surveys. The advantages of web-based surveys are numerous: They are less costly, less time-consuming to implement and more efficient (automated data collection) than other modes of data collection (Couper 2000). Moreover, web surveys allow for more interactive, visual or tailored formats, rather quick troubleshooting, automated scoring and reporting as well as access to larger samples. Web-based surveys have also been claimed to result in lower respondent errors and to lead to an increase in the completeness of responses. However, web surveys also come at a price: first and foremost, they depend on the availability of the Internet. Additionally, survey research suggests that response rates tend to be lower than those achieved with traditional modes such as mail-in surveys. And lastly, web-based surveys are less preferred by certain strata of society such as the elderly.

In the context of the ongoing post-referendum or election surveys conducted under the label FOKUS Aargau, this paper essentially attempts to answer the following three questions:

1. Does the Web-only mode affect the response rate, and thus the quality and size of the sample? Sample size is insofar important as the standard errors associated with the point estimates increase with lower sample size.
2. Does the Web-only mode affect the structure of the sample, namely its representativity? To make inferences from a sample about the population we require members of the target population to have the same chance of being selected.
3. Does the Web-only mode affect the quality of the answers? A further element of data quality is the rate of item non-response which should of course be as low as possible.

2 The FOKUS Aargau project

FOKUS Aargau is a Center for Democracy Studies Aarau (ZDA) project initiated in 2018, in which post referendum and election surveys are conducted for the Canton of Aargau (see more at fokus.ag). To date, two 2018 referendum votes, one in September (Higher Taxes for Millionaires) and the other in November (Forest Initiative, Voting Rights for the Swiss Abroad), as well as the Swiss general elections in October 2019 (for the National Council, the Council of States for the Aargau electorate, and the Cantonal Government) were covered(Milic et al. 2018, 2019a, 2019b, 2019c; Serdült et al. 2019). FOKUS Aargau uses address-based sampling (ABS) – the addresses are drawn from official register data of the Canton of Aargau – and postal mail for contacting sample members with a request to respond either using the Internet or by filling out the paper-based questionnaire. In other words, the FOKUS survey is a self-administered survey that informs the addressees by mail that they have been selected to participate in a research survey to which they can respond either

by using the paper questionnaire (a prepaid envelope is included in the letter) or by means of a Web questionnaire (accessing a link and using a login). The respondents themselves, therefore, choose the mode of response.

2.1 Procedure of the survey and experimental setup

In principle, the survey is conducted in the following way

1. The cantonal statistical office draws a random sample (full name, address and sex) of the Aargau electoral population on the basis of the population register.
2. The statistical office sends this sample directly (and in encrypted form) to the survey institute commissioned by the ZDA, which invites the sample population by postal mail to participate (online or by print questionnaire) and carries out the actual survey.
3. The survey institute *anonymizes the raw data* after the end of the survey (i.e. full name and address are deleted from the data set) and transmits them to us, the FOKUS Aargau team of the ZDA, which carries out the actual data analysis and prepares the report for the post-election analysis. This means that the ZDA never discovers the identity of the survey participants and cannot assign the results to individual persons. Even the commissioned survey institute cannot draw any conclusions about individual persons, since the relevant personal data has been irrevocably destroyed.

This procedure was defined in consultation with the Chancellery of the Canton of Aargau and was a condition for being allowed to use addresses from the population register. The entire procedure was also contractually defined. Data misuse or other violations of the contract would result in considerable financial penalties. The main issue was to avoid making individual level data from the survey identifiable. It was also not necessary to have access to the identity of the participants in order to the conduct of the FOKUS Aargau survey analyses and reporting.

In addition to providing post referendum or election reports, the FOKUS Aargau project also aims to explore special topics or methodological questions related to survey research in general. From the beginning of the project the quality of the samples generated by different modes of survey response was a concern. On the occasion of the October 2019 general elections in Switzerland, the FOKUS Aargau team at the ZDA therefore conducted an experiment evaluating how the response rate as well as the sample quality might differ in the case of an online versus an offline survey. Based on the official population register, 10'000 randomly drawn Swiss citizens who live in the Canton of Aargau received an invitation letter asking them to take part in the survey (no pre-notification letter was sent out). Whereas half of the sample received a paper questionnaire and were offered the opportunity to hand in their respective questionnaire by mail with a pre-paid envelope or online, the other half was only provided with a link (or QR-code) with which to access the online survey.

The general setup is visualized in Figure 1. The invitation letter went out a few days before the actual election day, which was the 20 October 2019. Swiss elections are traditionally on a Sunday.

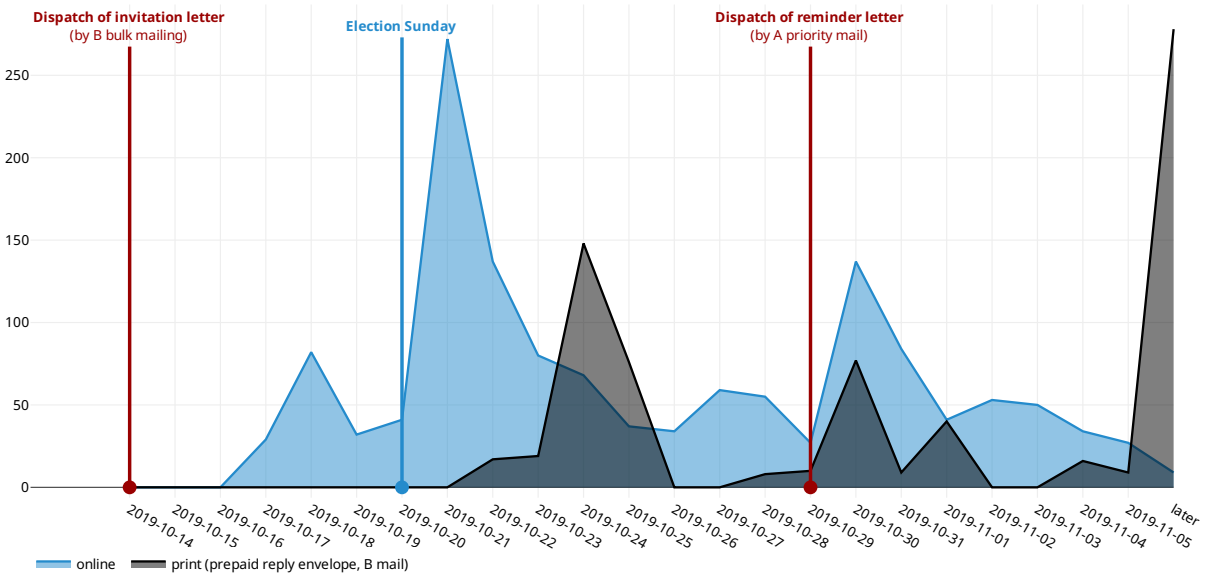

Figure 1: Completed interviews over time

Since most Swiss voters these days use the postal services to mail in their vote during the relatively long phase of three weeks prior to the elections, we had to strike a balance. Sending out the invitation letter too early would seem odd to the recipients, but too late, on the other hand, would risk losing a relatively large part of potential respondents voters for whom the elections would by then already be a thing of the past. Respondents choosing or having no choice but to use the online channel filled in questionnaires soon after having received the invitation, with questionnaires beginning to trickle in from 17 October 2019 and clearly peaking on the Monday after the election date, 21 October 2019, when all the results and media coverage were readily available. A second peak for the online mode can be observed right after the reminder letter went out on 29 October 2019. Pre-paid mail-in survey responses naturally arrive later, with the Swiss post usually delivering non-priority mail within 2-3 working days. Peaks can be observed at the beginning, after the reminder as well as on and after the official deadline given in the reminder letter. Whereas the larger peak (30 October 2019) for the mail respondents after having received the reminder is due to participants sending in the questionnaire before the given deadline, the second one (1 November 2019) is related to the reminder.

3 Response Rate

A high response rate within a drawn sample is a highly desirable goal. Dey (1997: 215) went so far as to consider a high response rate "akin to reaching nirvana." While we might not go quite so far, the advantages of a reasonably high response rate are obvious. The higher the response rate, the more likely we are to come up with unbiased sample estimates.[1] Furthermore, statistical power, for example for subgroup analysis, increases with larger samples as well.

Factors that affect participation in online surveys known, as shown in the literature, are a prenotification letter or card, the actual invitation to participate in the survey, the topic and sponsorship as well as incentives (Tourangeau et al. 2013: 43). Reviews of electronic survey research have suggested that response rates tend to be equal to or lower than those achieved with traditional modes such as mail surveys (Cook et al. 2000, for a review of factors affecting response rates for electronic surveys; e.g. Fricker and Schonlau 2002). Although findings specifically comparing mail surveys to e-mail/Web surveys are not completely consistent, with some evidence indicating similar (Truell et al. 2002) and some indicating higher response rates for e-mail/Web surveys (Cobanoglu et al. 2001; McCabe 2004), overall, there appears to be substantially more evidence supporting higher response rates for mail surveys (e.g., Cole 2005; Klassen and Jacobs 2001; Kwak and Radler 2002; Leece et al. 2004). Thus, evaluators planning to use a Web-based survey face the likely limitation of lower response rates relative to the more traditional approaches. Given this situation, to improve response rates many consider using a mixed-mode approach, in which another mode is used to contact those who do not respond to the initial e-mail/Web contacts (Roberts et al. 2016).

3.1 Response Rate: Comparison of survey methods

Response rates for the FOKUS Aargau surveys have to date been within the expected range. For most Swiss studies with similar survey designs, response rates are typically around 20-30 %. For the September 2018 referendum vote (a citizens' initiative on higher taxes for millionaires submitted by the young socialists) a relatively low 16.9 % of the 8'000 initially sent out requests to participate in the survey received a favorable response, resulting in a fully useable interview. The choice of the mode for filling in the interview questions was fifty-fifty: half online and half by postal mail. As for the reasons for the relatively low response rate we can only speculate. The referendum turnout certainly has an impact on response rates. Voters have a much higher probability of taking part in a post-election survey than non-voters. Possibly, the response rate for surveys in the Canton of Aargau might be low in general, regardless of what a survey is about. However, we are not aware of such an Aargau response rate bias from the literature, nor have we undertaken any further empirical investigations as yet to find out whether such a bias might exist. In addition, unlike many other surveys currently being carried out, we did not offer any special incentives (such as cash vouchers to be used in any of the major supermarkets). However, an alternative explanation could

[1] There is a long-standing and ongoing debate on how low response rates affect data quality. There seems to be a consensus among survey researchers that low response rates indicate *potential* bias. Nevertheless, there is no doubt about the fact that the *risk* of biased results increases with lower response rates.

be the fact that FOKUS Aargau surveys are as yet not well known to the general public, resulting in a certain skepticism towards an unfamiliar name, possibly also towards an institution (the Center for Democracy Studies Aarau – ZDA) which is also not very well known among the general public. The development of the response rate for the FOKUS Aargau surveys over time so far seems to support this explanation. Prominent press coverage might, on the other hand, have helped to generate a positive image and to make the name of the survey better known among the general public, allowing FOKUS Aargau to slowly create a brand of its own. For the November 2018 referendum votes (forest initiative, voting rights for the Swiss nationals living abroad) only two months later, and using 5'000 randomly chosen addresses among the Swiss population in the Canton of Aargau, a more impressive 23.4 percent responded favorably to their personal invitation letter. Again, the split between online and paper responses was almost fifty-fifty. For both of the 2018 FOKUS Aargau surveys the respondents were free to choose either of the two offered modes, that is, online or paper. For the data used in this paper, data resulting from the October 2019 general elections, sending out 10'000 survey invitations by mail (with one reminder, as in the previous cases) resulted in 2'146 valid cases, and thus a response rate of 21.6 percent. Since 5'000 out of the 10'000 invitation letters did not include the option to mail in the survey, and thus permitted only the electronic mode, the rate of online respondents was considerably higher this time, creating a split of almost two thirds online versus paper respondents (64.7 % versus 35.3 %).

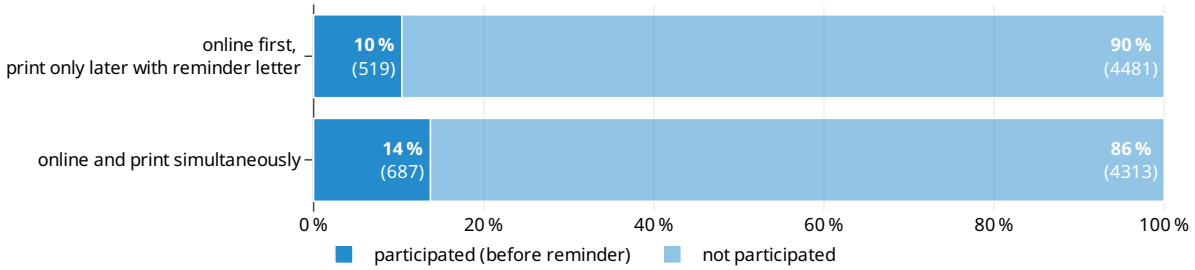

Figure 2: Survey participation by questionnaire provision type (before reminder letter)

When looking at mode-specific returns of survey responses before sending out the reminder letter 30 October 2019, the group which had the option of responding via paper or online shows a clearly higher response rate of 14 percent. In the group with potential respondents who were only offered the option to respond online, the response rate is at 10 percent (see Fig. 2). Results are consistent with the to date most encompassing experimental mixed-mode survey study for the Swiss context by Roberts et al. (2016: 27). A credible explanation for this pattern is given by Tourangeau et al. (2013: 49). Even for people inclined to participate in a post-election survey, a conflicting sense of obligation combined with something considered a nuisance might lead to procrastination on part of the respondents. For those who received a paper version of the questionnaire, the sheer fact of having a physical object waiting to be filled in might function as a reminder, eventually resulting in getting the job done. For obvious reasons, this is less the case for those with an invitation letter only.

All those in the sample who did not respond by a set date were sent a reminder via regular mail (cf. Fig. 1). For pragmatic reasons we refrained from expanding the study with the option of not

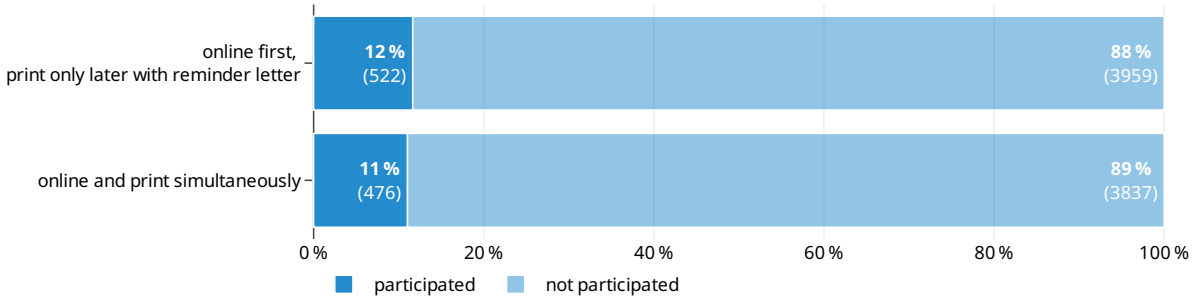

Figure 3: Survey participation by questionnaire provision type (after reminder letter)

sending a reminder letter to a part of the two groups of potential survey participants. Whereas not sending out a reminder letter to parts of our two sample groups would have improved the experimental design, it might have hurt the overall response rate. Having enough respondents, especially among respondents with a clear party identification, is crucial for such a post-election survey. The general importance of the reminder letter becomes apparent when looking at the response rates which were generated among the two sample groups after the letter was thought to have reached potential respondents (since the reminders were sent out by priority mail, arriving typically one day later). Among the ones still remaining without a response up to the date of the reminder letter, the response rates in both groups were almost equal (see Fig. 3), at 12 and 11 percent, respectively. Due to the limitation of our experimental design, we can, however, not be certain whether some of the survey responses arriving after the reminder would have been sent in any case, even without the reminder letter.

4 Representativeness

4.1 General remarks

Whatever sampling approach one chooses, the goal (normally) is to make (externally valid or accurate) inferences about some broader population. When using probability-based samples, the validity of the inferential claims depends on whether the requirements of a probability-based sample are met or not. The most important requirement of probability sampling is that every member of the target population has a known, nonzero chance of selection. When these conditions are satisfied, the sample should, in all probability, resemble the target population with regard to most if not all demographic or political variables.[2] If, however, the above-mentioned provisions are not met – which it is almost always for various reasons (e.g. coverage error, non-response[3], etc.) – the sample's representativeness might be (heavily) compromised. Thus, the quality of surveys can to a certain degree be measured through their representativeness (Cornesse and Bosnjak 2018; Schouten et al. 2008). Representativeness, in turn, refers to how well the sample drawn compares with (or is representative of) the population at the focus of a study.[4] Thus, the question addressed in this section is: To what extent are the online and print samples representative regarding certain variables of interest? Moreover, does providing only a Web questionnaire hamper the representativeness of the sample to a significant degree?

There are essentially two approaches employed in the relevant literature to measure the concept of representativeness (Cornesse and Bosnjak 2018): The first and more straightforward approach is to simply compare survey responses with known population data. Official population statistics can almost always be considered as a "benchmark" reflecting the true values of the population at the focus of the study (e.g. Keeter et al. 2000). The second approach (also called "record-linking") compares the sample estimates to those of the gross sample. Examples of such measures in the field of survey research include the so-called R-Indicators (Schouten et al. 2008) or balance and distance measures (Lundquist and Särndal 2013). The downside of this second approach is, however, that only those characteristics can be assessed that are also included in the gross sample. Quite often, only a few variables are available – and more often than not these few variables are only weak predictors of the variable that is of interest to the researcher. Furthermore, depending on how the gross sample was drawn (and, particularly, from which list), it might already be biased compared to the larger target population. In our case, the gross sample was drawn from a (almost) complete list of all members of the target population (Cantonal register). However, the only information

[2] Compare e.g. the structure of the gross sample (i.e. the units that are sampled in a first step) drawn from Cantonal register data with the Cantonal population statistics in Table 1.

[3] Since the FOKUS Aargau random sample was drawn from the Cantonal population register, non-coverage by definition cannot be a source of survey error. However, non-response (amounting to roughly 78 percent for the third wave of FOKUS Aargau) might very well be a source of error (Betlehem et al. 2011).

[4] Representativeness actually is a vague term with a multitude of possible meanings (see Kruskal and Mosteller 1979). Some researchers even reject the term entirely (Schnell 1993). Others use different terms like "representativity" (Betlehem et al. 2011). Kruskal and Mosteller (1979) define a representative sample as a sample, where the target individuals were chosen randomly and entirely by chance (i.e. a probability-based sample), thereby referring to the process of generating a sample rather than to its structure. From the perspective of total survey error (Groves and Lyberg 2010) survey (non-)representativeness includes all sources of errors of nonobservation: Coverage error, sampling error, nonresponse error and adjustment error (Groves and Lyberg 2010; R. M. Groves et al. 2004).

available for *all units of the gross sample* are gender, age, marital status and household size. Thus, the representativeness assessment will be conducted by comparing the estimates of the different net samples to those stemming either from population register data or from the gross sample.

4.2 Does weighting solve all of the representativeness problems?

A remark that is apt here is that one way of dealing with lack of representativeness is to weight the sample. Indeed, weighting procedures[5] are meanwhile commonly applied in all kind of surveys to compensate for nonsampling errors (for an overview see Brick (2013) and Kalton and Flores-Cervantes (2003)). From this perspective, one could argue that most if not all samples – including probability samples – are biased (due to high rates of non-response, see e.g. Brick and Williams (2013)) and the real concern should therefore not be with drawing a sample that is as representative as possible, but rather to weight it properly.[6] While there certainly is truth in this claim, it would be wrong to say that the representativeness of a sample is irrelevant. In addition, due to the fact that a highly representative sample is always better to start with than a biased sample, all weighting methods are based on either explicit or implicit assumptions about the nature of non-response (Brick 2013; Franco et al. 2017; Kalton and Flores-Cervantes 2003).[7]

As a result, weighted estimates "have more desirable properties only if the assumptions underlying the weighting procedure are satisfied" (Franco et al. 2017: 1). When, however, in the real world those assumptions made in applying adjustments are violated (e.g. sampling probability is not only different between sub-groups, but also within those sub-groups), the weighted estimates might not reduce or at least not totally eliminate coverage or non-response bias (Chang and Krosnick 2009; see e.g. Dever et al. 2008). As Tourangeau et al. (2013: 33) report in their meta-analysis of the efficiency of weighting adjustments, "adjustments remove only part of the bias, at most around three-fifths" and "sometimes increase the biases relative to unadjusted estimates,

[5] Our line of argumentation is general and pertains to any weighting method (post-stratification, inverse-probability weighting, propensity-score weighting, MRP or raking). For the sake of simplicity, we use the word "weighting" to refer to all such methods.

[6] For an excellent overview over the ongoing debate in the survey research literature about whether and when probability and nonprobability sample surveys produce accurate estimates of a larger population see Cornesse et al. (2020).

[7] Weighting procedures are based on certain assumptions regarding nonresponse. A first assumption is that survey participation is not a predetermined quantity (Groves and Couper 1998), but based on response propensities (Rosenbaum and Rubin 1983). This stochastic view assumes that individuals have a specific, but usually unknown probability of survey participation. Starting from this assumption, further assumptions about the randomness or nonrandomness of non-response can henceforth be formulated: A first assumption is that non-response in interviews is completely random (Missing Completely at Random, "MCAR"). If this assumption were true, no sample correction would be necessary. For, in such cases, the missing values are a random sample of all values and not related to any observed or unobserved variable. Thus, the data will not be biased. However, this assumption is – based on everything we currently know about non-response research – highly unrealistic (Groves and Couper 1998). A second assumption (Missing at Random, "MAR") acknowledges that unit non-response is not random, but that it can be fully accounted for by other variables (auxiliary variables) collected in the survey (such as gender, age, etc.). Thus, missingness depends only on observed characteristics and not on the missing values themselves. The last assumption – Missing Not at Random, ("MNAR") – is the most "pessimistic" and the most difficult to handle (for an overview see Little and Rubin (2019)): In a nutshell, MNAR means that there is a relationship between the propensity of a value to be missing and its values. If the data is MNAR, additional assumptions have to be made in order to model the missing data mechanism. Whether those assumptions hold true or not, however, is usually not verifiable. Thus, in summary, there is often no possibility to empirically test these assumptions for a particular data set or to properly uncover MNAR. In most cases – for pragmatic rather than theoretical reasons – MAR is assumed. Whatever assumptions we make, if they do not hold true, the weighting adjustment might not be effective in reducing nonresponse bias. As a matter of fact, in some extreme cases, there might be a situation where the weighted results are even less preferable than the unweighted ones.

sometimes by factors larger than 2." Several other researchers also found that weighting on demographic variables is rather unlikely to solve all of the representativeness problems (Cornesse et al. 2020; Couper et al. 2007; Rookey et al. 2008; Zhang et al. 2008). Thus, even if weighting is to be applied, it makes sense to assess the representativeness of a sample in order to have an idea of how much one depends on these additional assumptions required when employing weighting adjustments.

4.3 What do we know about the representativeness of online samples?

There is a wealthy and ever-growing body of literature dedicated to the topic of the representativity of Web surveys (Bandilla et al. 2009; Bieber and Bytzek 2012; Chang and Krosnick 2009; Cornesse and Bosnjak 2018; Grewenig et al. 2018; Hoogendorn and Daalmans 2009; Kohut et al. 2012; Szolnoki and Hoffmann 2013; Yeager et al. 2011). However, most of these studies compare *non-probability-based* Web surveys to other, *probability-based* methods of data collection. Not surprisingly, there are considerable differences reported in these studies. But so far, there are only a few studies assessing the representativeness of *probability-based* Web samples (Blom et al. 2016). Nevertheless, we should expect the same characteristics to be overrepresented in probability-based as well as in non-probability-based Web surveys – albeit to a (much) lesser degree. Because, clearly, the structure of a Web sample will be heavily influenced, firstly, by the fact whether the respondents have Internet access or not, and, secondly, to a lesser degree by the respondents' web affinity. In short, the digital divide (Norris 2001) will also be found in the structure of the Web sample.

There is a great deal of evidence to the effect that the "digital" population does indeed differ from the "non-digital" or "offline" population and that it does so in a number of ways. The rates of Internet coverage not only differ across demographic groups but, and more importantly for research on political attitudes, also across attitudes of groups. As for the demographic differences, the digital divide follows along the well-known lines of age, gender, income and education: Young people, people with a high level of formal education and comparatively high income are mostly overrepresented in Web surveys. In addition, men are often overrepresented (Bandilla et al. 2009; Bieber and Bytzek 2012; Blom et al. 2016; Hoogendorn and Daalmans 2009; Szolnoki and Hoffmann 2013; Yeager et al. 2011). Thus, the structure of Web surveys is to some degree the reverse mirror of the structure of telephone surveys: in the latter, young, highly educated men are among the groups most difficult to reach, whereas in Web surveys they are, as already mentioned, clearly overrepresented. Elderly women, on the other hand, are less likely to participate in Web surveys than in the more traditional survey modes such as telephone surveys (Hawkins et al. 2013; Kwak and Radler 2002). When considering online polling about political topics in particular, the results are somewhat less clear (Bieber and Bytzek 2012; cf. e.g. Faas and Schoen 2009). Participation in elections or votes is usually overestimated in Web surveys – but equally so in "offline" surveys (Selb and Munzert 2013). As a matter of fact, studies suggest that people are less likely to be influenced by social desirability effects – and participating in an election is still considered to be a norm – with regard to online surveys than to other modes (Holbrook and Krosnick 2010; Kreuter et al. 2009). As for political attitudes, in the US, Web participation relates to some extent to liberal attitudes and

higher socioeconomic status (Best and Krueger 2005). There is, however, generally little research on the representativeness of Web surveys conducted in Switzerland (Roberts et al. 2016).

4.4 Representativeness of the online sample in demographic and political terms

In a first step, we compared the "raw" (i.e. unweighted) demographics of the online and print sample with population statistics from the Cantonal register (Kantonales Einwohnerregister) and with the gross sample (table 1), which is a simple random sample (without replacement) drawn by the Statistics Bureau of Aargau from the sampling frame (Cantonal register). In a second step, we compared variables of interest between the general online sample and the sample of those respondents who received only a Web login (but no paper questionnaire). The latter could not choose between the Web and the paper version of the questionnaire. In essence, by comparing these two groups, we want to find out whether target respondents who prefer the paper variant, but are not provided with it, refrain from participating in the survey, or not. To test the significance of the differences, the estimates in the samples (mail, Web, Web only) were compared using a two-sample z-test comparing two proportions. We took the population's proportions into account where these quantities were known from the population statistics.[8]

Our findings are mostly in line with other, similar studies: Participants of Web surveys are on average significantly younger than those of more traditional modes like postal surveys. They also tend to be more educated and more likely to be male. All these differences between the online and postal sample are significant and they also stand up to a multivariate test (see figures 5–7 in the Appendix for the corresponding effect plots). Furthermore, there is no clear pattern that allows us to say which sample has a larger bias. The print (or mail) sample is more representative with regards to gender, older age groups and low education. The online sample is closer to the population benchmarks when it comes to younger and more highly educated people. However, and more importantly against the backdrop of our rather practical research question, the combined sample ("Mixed-mode-sample") clearly has a higher representativity than both, the online and the mail sample alone. The Web only sample differs only slightly from the general Web sample and it also strongly resembles the sample of "digital" respondents who *had* a choice between the Web and the paper version of the questionnaire (results not in table 1).

As for political behavior and preferences, the comparison between the Web and mail sample yields

[8] The test statistic for testing the differences in the proportions between Web and mail sample when the true variance is known (gender, age groups) is:

$$z = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) - (p_1 - p_2)}{\sqrt{\left(\frac{x_1+x_2}{n_1+n_2}\right)\left(1 - \frac{x_1+x_2}{n_1+n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The test statistic for testing the differences in the proportions between Web and mail sample when the true variance is unknown is:

$$z = \frac{\left(\frac{x_1}{n_1} - \frac{x_2}{n_2}\right) - 0}{\sqrt{\left(\frac{x_1+x_2}{n_1+n_2}\right)\left(1 - \frac{x_1+x_2}{n_1+n_2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Table 1: Representativeness of Web and mail sample in terms of demographic characteristics (proportions in %)

Variables	census	gross (n=10'000)	whole (n=2'204)	mail (n=765)	Web (n=1'439)	Web only (n=515)
gender					3.376^{***}	0.109
male	48.45	48.28	50.0	45.4	53.0	52.6
female	51.45	51.72	50.0	54.4	47.0	47.4
age					16.248^{***}	5.108^{***}
20–29 years	14.47	16.65	14.1	4.64	16.38	17.1
30–39 years	14.76	14.76	11.8	7.81	14.25	10.7
40–49 years	15.71	14.85	15.2	9.01	18.95	14.6
50–59 years	20.10	19.61	20.8	18.68	22.65	24.7
60–69 years	16.21	15.69	18.0	23.71	15.53	18.5
over 70 years	18.74	18.44	20.2	36.16	12.25	14.6
education					11.225^{***}	1.462
obligatory level	18.1	-	5.6	10.4	3.0	3.1
secondary level	49.9	-	47.0	53.1	43.7	43.7
tertiary level	32.0	-	47.4	36.6	53.3	53.2
marital status					10.149^{***}	2.814^{*}
single	28.0	30.8	24.1	14.5	29.2	26.6
married	56.3	52.5	62.3	65.1	60.9	61.6
widowed	5.6	5.7	3.7	7.6	1.7	2.7
divorced	10.1	10.9	9.9	12.8	8.3	9.1

Sources: Kantonales Einwohnerregister 2018 (census data for gender and age), Strukturerhebung 2019 (education). The numbers for marital status are from the cantonal statistical office of Aargau and are estimates. Education: only for >25 years. Notes: The z-statistic (p-value) for the respective two sample z-tests are reported for the comparison between the web and the mail sample as well as between the Web only and the Web with choice samples.

ambiguous results. In both samples, there was a considerable and, more importantly, an equal amount of overreporting regarding participation in the election. While in reality, only 45 percent of the Argovian electorate turned out to vote, roughly 80 percent of respondents in both samples reported to have cast a valid vote on 20 October 2019.[9] It has been long recognized (see in particular the Swiss studies from Sciarini and Goldberg (2016) and Hugi (2014), who used official voter records to verify who voted), that it is not only overreporting in the sense of intentional misreporting owing to social desirability, but – to a large extent – also sampling errors like non-response[10] and non-coverage which contribute significantly to overestimations of turnout (Selb and Munzert 2013). It is safe to assume that two of these three categories of errors (misreporting and non-response) affected the self-reported turnout in the FOKUS Aargau survey, too. It also does not come as much of a surprise that there are no differences in the aggregated self-reported turnout between the Web and the mail sample. Both surveys are self-administered and thus – due to the self-completion format in which interviewers are not present (e.g. Kreuter et al. 2009) – should in equal measure be less subject to social desirability effects than other modes. But precisely the latter claim – to show less of a likelihood of this effect than other modes – is not corroborated by the FOKUS Aargau results. The overrepresentation of almost 35 percent is similarly high or even higher than for other "offline" studies (e.g. VOTO or SELECTS).

[9] On 20 October 2019, three elections took place in the Canton of Aargau: The elections to the National Council, the elections to the Council of States and, lastly, an election for a substitute to the Cantonal government. We asked the respondents whether they took part in one of these elections. If we subtract the number of those who did not cast a valid vote in the elections to the National Council, the gap between reported and actual participation reduces (slightly) to 30 percentage points.

[10] Voters are almost certain more likely than nonvoters to participate in postelection surveys. Thus, responding to the surveys will correlate with the propensity to vote.

Table 2: Representativeness of online and print sample in terms of political characteristics (proportions in %)

Characteristic	**census**	**whole** (n=2'204)	**mail** (n=765)	**Web** (n=1'439)	**Web only** (n=515)
participation				0.046	3.768***
participants	45.4	80.0	80.0	79.9	85.2
non-participants	54.6	20.0	20.0	20.1	14.8
party vote			6.942***	2.360	
SVP	31.5	24.3	33.8	19.4	19.2
SP	16.5	20.1	17.9	21.2	22.5
FDP	13.6	14.9	14.5	15.1	15.3
CVP	9.9	9.1	8.4	9.5	10.0
GPS	9.8	12.1	10.7	12.8	13.0
glp	8.5	11.2	8.2	12.8	11.3
BDP	3.1	2.1	1.4	2.5	2.6
EVP	3.6	3.7	3.1	4.0	3.9
EDU	1.0	0.8	0.4	1.0	1.4
others	2.5	1.7	1.7	1.7	0.9

Notes: The participation rate refers only to Argovians eligible to vote and living in the Canton of Aargau. Argovians living abroad ("Auslandschweizer") are not included.

Looking at the political preferences (i.e. party vote in the elections to the National Council), however, we find stark differences between the Web and the mail sample. "Onliners" tend to vote disproportionally more often for left parties, while the mail respondents more likely voted for the SVP. But, as the results of multivariate logistic regression of the determinants of mode choice (see table 4 in the Appendix) suggest, this is mainly due to sociodemographic characteristics which differ strongly between Argovian party electorates: SVP and CVP voters are on average significantly older than voters of left parties, particularly voters of the Greens and the glp party (Milic et al. 2019c). Party preference (measured through PID) for the most part had no significant impact on the mode preference when holding age, gender, and education constant, and other controls. Thus, as could be expected, the mode differences are most likely not caused by political preferences but rather are a by-product of the digital divide. However, one particular attitude had a fairly strong effect on mode preference: The attitude towards a nationwide introduction of e-voting. This attitude, in turn, is strongly influenced by Web-affinity, Web-familiarity, but also the general trust/distrust towards digital processes of all kind (e-government, e-banking, e-voting) (Milic et al. 2016). Thus, there is a significant risk of "scaring off" a substantial part of the population by going fully online.

In summary, consistent with other studies, we find that the exclusion of offline households produces significant coverage biases, while the inclusion of these households in the sample improves the representativeness of the survey significantly (Blom et al. 2016). Whether that bias can be fully compensated by employing weighting adjustments remains open.

5 Survey response quality

5.1 Item non-response

The completeness of answers or item-response rate is another dimension of data quality (Hoonakker and Carayon 2009). Item non-response refers to the situation where data are missing for some (but not all) items or questions.[11] The level of item non-response not only depends on the survey mode, but on numerous other factors: on characteristics of the participants (level of motivation, interest, commitment, etc.), the wording of the question and certain item or survey design factors (Tourangeau et al. 2013). With design features or choices, we mean e.g. paging vs. scrolling designs and, closely associated with it, the question whether "forcing" elements (i.e. a forced answering option) were incorporated or not. Lastly, it is almost a truism to say that the absence of a "don't know"-category increases item-dropout. Precisely because of the manifold varieties of design choices in existing Web surveys, it is not very surprising that the findings on item non-response in Web surveys are highly inconsistent. Some studies reported significantly higher rates of item non-response for mail surveys than for Web surveys (Denscombe 2009; Kwak and Radler 2002; Tourangeau et al. 1997), while other studies found the complete opposite (Denniston et al. 2010; Manfreda et al. 2008). And indeed, one study (Wolfe et al. 2009) found no differences at all between Web and mail surveys. As mentioned before, the reason why all these studies came to such different conclusions, has most likely to do with different design choices and features of the Web surveys.

In order to assess the item non-response rates of the FOKUS Aargau Web and mail samples properly, it is necessary to briefly sketch the features of the FOKUS Web survey. There was one key question in the FOKUS Web survey which required an answer (i.e. had a forced answer format): The question whether one took part in the election or not. Thus, respondents could not skip that particular question, but were forced to give an answer. All other questions were non-mandatory, i.e. they were not required to be answered in order for respondents to proceed to the next question. However, if a respondent did not answer a question, a window popped up asking whether not answering the question at all was intentional. If so, the respondent subsequently could tick a corresponding box in that pop-up window. Furthermore, with only a few exceptions (marital status, denomination, age, gender), all questions included a "don't know" option. Clearly, requiring respondents to answer every question or – as in our case – to be reminded to do so, is a long-established way to eliminate item missing data. It comes with a price though: increased breakoffs. Whether our design increased breakoffs or not, cannot be answered, since we did not conduct an A/B-test with different Web features. In total, roughly three percent of those taking the Web questionnaire started the interview but dropped off at a later stage without completing the interview.

[11] Another survey quality measure are breakoffs (also called abandonments or terminations). They affect Web surveys, whether they are probability or non-probability samples. Clearly, breakoffs are much more common in self-administered surveys (like Web surveys) than in interviewer-administered modes (like telephone surveys, see Tourangeau et al. 2013). With online data, breakoffs can be detected and quantified precisely. However, in our case, both modes are self-administered. Moreover, there is no way of comparing breakoff rates between Web and mail samples. And the reason for this is rather obvious: The extent of breakoffs in mail surveys is simply not known. We don't know how many respondents started to fill out the questionnaire, but failed to complete it or – perhaps – simply forgot to send it in.

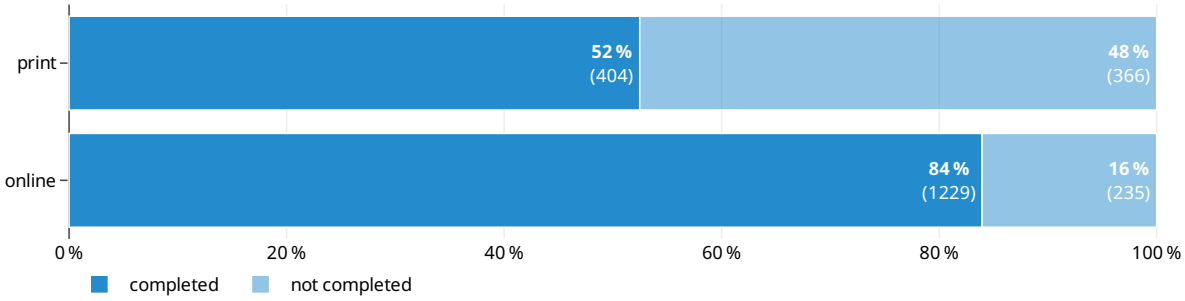

Figure 4: Completeness of the interview by questionnaire channel

If we compare the proportions of those completing the interview without skipping a single question[12] (see figure 4) between the Web and the mail sample[13], we find stark differences: The completion rate (i.e. completeness of all questions) for the Web sample amounts to roughly 84 percent, whereas the completion rate for the print sample is a (rather meagre) 52 percent. How can this large difference be explained? First, the paper questionnaire includes a bunch of filter questions and it might well be that some respondents simply could not follow the corresponding logic, unintentionally skipping a few questions. Secondly, when comparing the incompletion rate between the different question blocks (see table 3), we find particularly large differences between the Web and the mail sample in block 3 and block 6. Block 3 included mostly questions testing the factual political knowledge of participants and block 6 included the probably most challenging filter logic within the paper questionnaire.

Table 3: The completeness rate of question blocks for the Web and the mail sample

	survey mode		
Block	**mail**	**online**	**z-test (p-value)**
block 1 (sociodemographics)	98.7	98.3	0.739
block 2 (basic political questions)	99.6	98.0	3.090***
block 3 (mostly knowledge questions)	84.9	96.1	9.388***
block 4 (political behavior)	99.1	95.4	4.603***
block 5 (media use)	91.6	94.6	2.785***
block 6 (political attitudes)	79.9	94.6	10.807***
block 7 (Big Five)	88.4	94.0	4.613***
block 8 (sociodemographics)	74.9	84.7	5.629***

To explain the *completeness* (or rather incompleteness) of the survey we estimated a logistic regression model with all the demographic characteristics of respondents which were available in the gross sample as well as the survey mode and some additional political variables of interest. As a matter of fact, in order to avoid tautology issues and, consequently, estimation problems, we calculated the number of missing answers to all questions in the blocks 3 to 8. This way we could include political interest and participation (questions from the first and second block respectively) as additional control variables. The dependent numerical variable (number of skips for all questions from block 3 to 8) was then dichotomized, whereby those respondents filling out all questions from the respective blocks were assigned the value 0, while the rest was assigned the value 1. Moreover,

[12] The Item non-response variable was constructed as a binary variable: 1=nonresponse, and 0=response.

[13] Only questions posed to all survey participants were included in this analysis.

we tested a second model, where the dependent variable consisted of missing answers to all questions except block 7. Block 7 includes questions referring to the so called "Big Five," enabling us to test whether questionnaire mode is significant even in the case of holding constant fundamental character traits. We further excluded "don't know" answers from the analysis.

	Dependent variable:	
	completeness (blocks 3-8)	completeness (exc. block 7)
	(1)	(2)
questionnaire mode (1=online)	−1.618***	−0.885***
	(0.116)	(0.114)
age	0.011**	−0.023***
	(0.004)	(0.004)
gender (1=female)	0.368***	0.544***
	(0.110)	(0.102)
household size	0.085	0.004
	(0.052)	(0.049)
married	−0.475***	−0.120
	(0.162)	(0.151)
divorced	−0.422*	0.028
	(0.218)	(0.208)
widowed	0.004	0.050
	(0.299)	(0.323)
participation	−0.240	−1.607***
	(0.151)	(0.134)
political interest	−0.184**	
	(0.083)	
trust in good		−0.114**
		(0.050)
nervous/insecure		0.192***
		(0.049)
Constant	−0.493	1.788***
	(0.405)	(0.383)
Observations	2,183	2,032
Log Likelihood	−1,077.138	−1,174.024
Akaike Inf. Crit.	2,174.276	2,370.048
Note:		*p<0.1; **p<0.05; ***p<0.01

The analysis shows that the questionnaire mode remains a highly significant predictor of item nonresponse even when holding constant age, gender as well as political interest and participation. As a matter of fact, the survey mode is the strongest predictor. In other words, the online group shows a significantly higher probability than mail participants of completing the interview without skipping any questions at all. Additionally, women and younger participants tend to skip questions more often than men and elderly survey takers. Not surprisingly, political interest also has an impact on the completeness of a survey. Politically highly involved participants have, firstly, a stronger incentive to answer the questions to their best knowledge and, secondly, are probably better equipped in terms of political knowledge to answer the questions. The results from model 2 render similar conclusions. The questionnaire mode remains one of the strongest predictors of

item non-response. In the absence of political interest (which we could not include into the model due to tautology issues), it is the fact whether one participated or not in the elections that predicts item non-response. Participation and political interest are highly correlated, and therefore, it is not surprising that participation exerted a significant effect on item non-response. In addition, two items of the "Big Five"-battery turned out to have a significant impact on the dependent variable: First, respondents who often feel insecure and nervous tend to skip questions more often than others and, secondly, respondents who have faith and trust in people were more likely to complete the survey without leaving any questions unanswered.

6 Conclusion

Election research, and political behavior studies in particular, still largely have to rely on survey data. In the political realm in particular the ramifications of conclusions drawn on candidates, governments and political parties, based on survey studies, can be immediate and consequential. Hence, special care must be taken by pollsters and political science researchers alike already at the very beginning of an analysis, when samples are drawn and design choices for a survey study are made. A high-quality sample is therefore fundamental and key for data analysis leading to an election report or, in the case of Switzerland, a referendum report. With three to four general referendum voting days per year in Switzerland, during which major topics are decided by the electorate on the national, cantonal and local level, post-referendum surveys are of particular importance. Such reports are well-established nowadays and play a part in the political discourse among citizens, political elite and media. However, perhaps due to the very high number of referendum votes and a limited institutional as well as financial support for post-election and post-referendum studies in Switzerland, there is still a lack of methodological research, in particular with regard to the quality of survey samples and different survey modes. The FOKUS Aargau series of post-referendum surveys and reports would like to make a contribution in that regard. In the present study we have mainly set out to investigate the influence of the survey mode on the quality of the samples, comparing respondents forced to choose the online channel versus the ones having the choice between online and the more traditional pen and paper version to be submitted by mail.

Our results can be summarized as follows. The initial as well as total response rate among potential survey respondents who had the choice of either using the paper version and sending it in via pre-paid mail or using the online version was somewhat higher than for the sample in which the only available channel to respond was the Internet. However, the difference in response rates between the two modes is not large and one might be led to conclude that the paper mode is more or less obsolete. However, response rates are not the only criterion when evaluating the quality of a survey sample. Regarding representativeness, the two sub-samples tend to complement each other. The sample with the mail-in option is more representative of a particular gender, older age groups and people with a lower level of formal education. For the online sample the pattern is the reverse. In total, the mixed-mode samples in combination clearly yielded the most representative sample. Regarding over-reporting, there is no difference between the two survey modes. The

most striking differences between respondents' answering mode are revealed when looking at the, for a post-election survey crucial, political preferences. Mail-in survey respondents clearly tend to favor identifying with and voting for the SVP, while online mode respondents more often vote for left-wing political parties. However, when controlling for other variables in a multi-variate model, party-based differences almost disappear and in their place a pattern of digital divide emerges, e.g. respondents favoring the nationwide generalization of internet voting tend to opt for the online survey mode. In sum, consistent with other studies, we find that the exclusion of offline households produces significant coverage biases, while the inclusion of these households in the sample improves the representativeness of the survey significantly. Non-response as the last of the three quality criteria compared among the two survey modes showed stark differences. With a meagre completion rate of 52 percent, the paper version fares much worse than the online mode for this survey (84 % completion).

In sum, we provisionally conclude that in the current environment the two modes in the FOKUS Aargau survey design attract significantly different strata of society into the two samples of our experiment. Hence, the two modes tend to complement each other and should, thus, both be kept. The price for completely replacing more traditional surveys such as mail-in solutions with an online version is still too high, in particular regarding the criterion of representativeness of the survey respondents. However, major advantages of the online survey such as the quick response (and therefore availability of data), high enough response rate and superior completion rate should not be overlooked and speak in favor of this mode.

Appendix

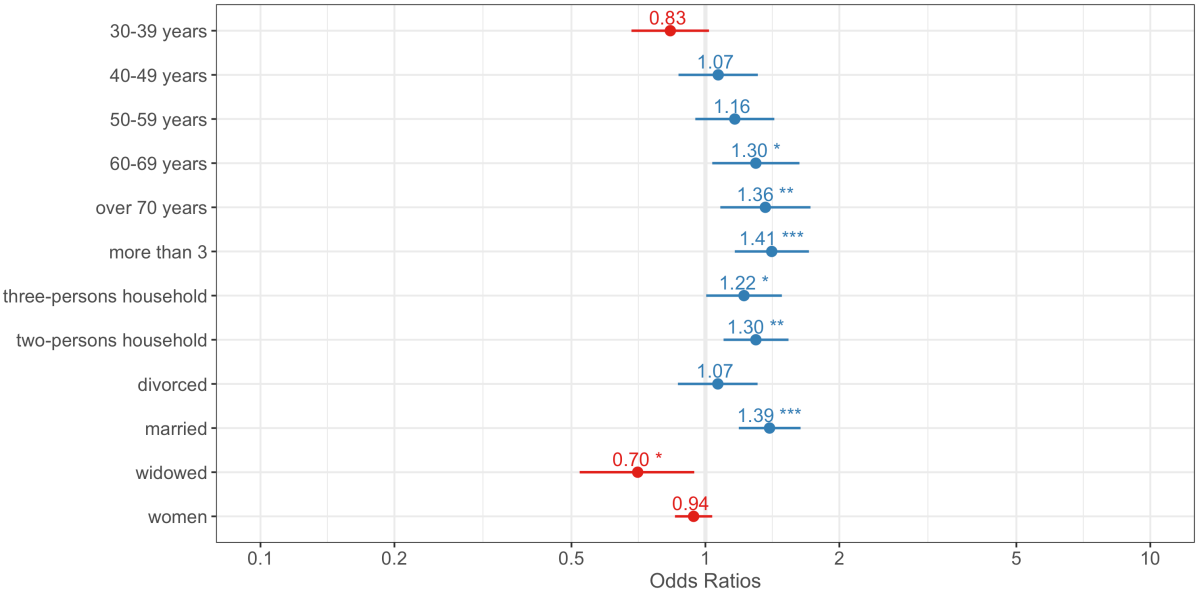

Figure 5: Determinants of survey participation, odds ratios, logistic regression (n=10'000)

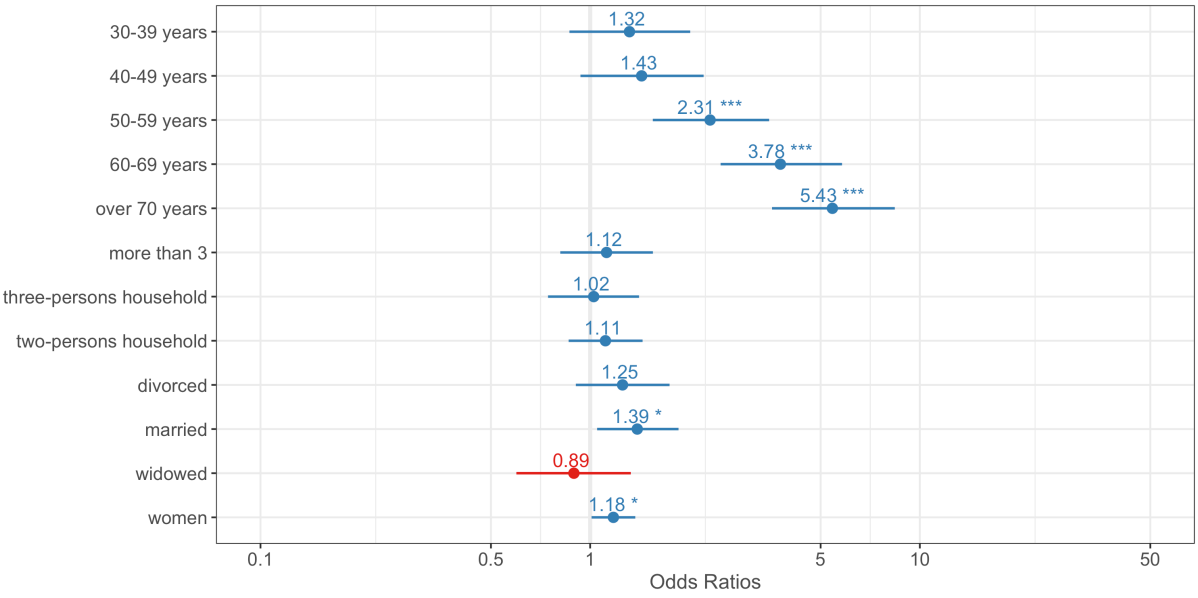

Figure 6: Determinants of print participation, odds ratios, logistic regression (n=10'000)

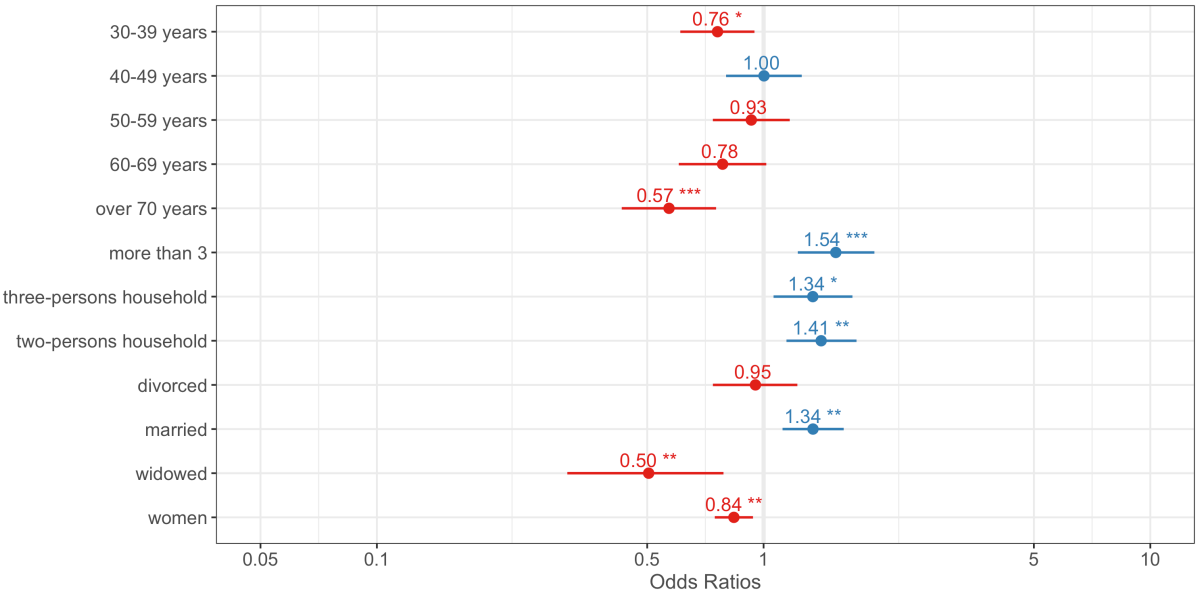

Figure 7: Determinants of online participation, odds ratios, logistic regression (n=10'000)

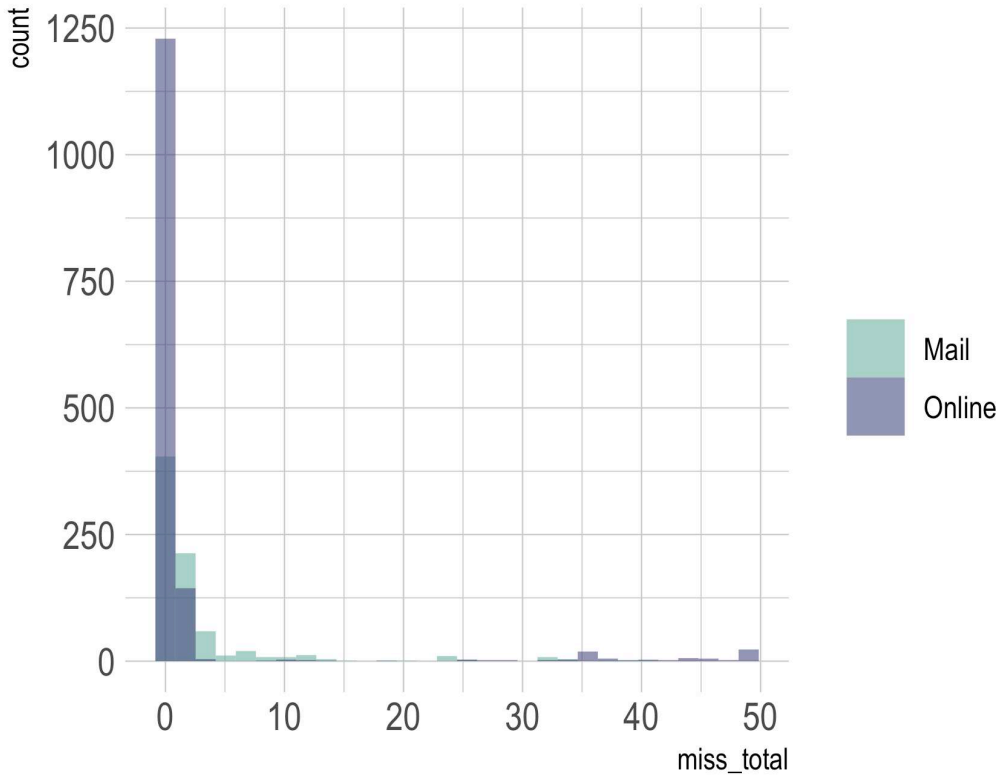

Figure 8: Number of missing answers for Web and mail sample

Table 4: Determinants of mode choice, logit coefficients (only respondents who had a choice between Web and mail)

	Dependent variable
	mode choice (1=online)
gender (1=female)	−0.203
	(0.131)
age group	−0.475***
	(0.045)
political interest (4-point scale)	0.317***
	(0.105)
education (3-point scale)	0.331***
	(0.076)
attitude e-voting (4-point Likert)	0.513***
	(0.062)
trust swiss democracy	0.039
	(0.032)
left-right self-pos.	0.053
	(0.076)
no PID	1.005***
	(0.271)
sp	0.663**
	(0.276)
fdp	0.282
	(0.222)
cvp	0.414
	(0.274)
greens	0.153
	(0.285)
glp	0.386
	(0.281)
other	0.535*
	(0.321)
Constant	−1.676***
	(0.610)
Observations	1,350
Log Likelihood	−743.460
Akaike Inf. Crit.	1,516.900
Note:	*p<0.1; **p<0.05; ***p<0.01

Bibliography

Bandilla, W., L. Kaczmirek, M. Blohm, and W. Neubarth (2009). Coverage- Und Nonresponse-Effekte Bei Online-Bevölkerungsumfragen. In Jackob, N. et al. (eds.), *Sozialforschung Im Internet. Methodologie Und Praxis Der Online- Befragung*. Wiesbaden: VS Verlag für Sozialwissenschaften (129–45)

Best, S. J. and B. S. Krueger (2005). Analyzing the Representativeness of Internet Political Participation. *Political Behavior* 27(2): 183–216

Betlehem, J., F. Cobben, and B. Schouten (2011). *Handbook in Nonresponse in Household Surveys*. New York: Wiley

Bieber, I. E. and E. Bytzek (2012). Online-Umfragen: Eine geeignete Erhebungsmethode für die Wahlforschung? Ein Vergleich unterschiedlicher Befragungsmodi am Beispiel der Bundestagswahl 2009. *Methoden, Daten, Analysen (mda)*: 185–211

Blom, A. et al. (2016). A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe. *Social Science Computer Review* 34(1): 8–25

Blumberg, S. J. and J. V. Luke (2007). Coverage Bias in Traditional Telephone Surveys of Low-Income and Young Adults. *Public Opinion Quarterly* 71(5): 734–49

Brick, J. M. (2013). Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics* 29(3): 329–53

Brick, J. M. and D. Williams (2013). Explaining Rising Nonresponse Rates in Cross-Sectional Surveys. *The ANNALS of the American Academy of Political and Social Science* 645(1): 36–59

Chang, L. and J. A. Krosnick (2009). National Surveys via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly* 73(4): 641–78

Cobanoglu, C., P. J. Moreo, and B. Warde (2001). A Comparison of Mail, Fax and Web-Based Survey Methods. *International Journal of Market Research* 43(4): 1–5

Cole, S. T. (2005). Comparing Mail and Web-Based Survey Distribution Methods: Results of Surveys to Leisure Travel Retailers. *Journal of Travel Research* 43(4): 422–30

Cook, C., F. Heath, and R. L. Thompson (2000). A Meta-Analysis of Response Rates in Web-or Internet-Based Surveys. *Educational and Psychological Measurement* 60(6): 821–36

Cornesse, C. et al. (2020). A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. *Journal of Survey Statistics and Methodology*(smz041). https://doi.org/10.1093/jssam/smz041 [accessed 28.1.2020]

Cornesse, C. and M. Bosnjak (2018). Is There an Association Between Survey Characteristics and Representativeness? A Meta-Analysis. *Survey Research Methods* 12(1)

Couper, M. P. (2000). Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*(64): 464–94

Couper, M. P., A. Kapteyn, M. Schonlau, and J. Winter (2007). Noncoverage and Nonresponse in an Internet Survey. *Social Science Research*(1): 131–48

Denniston, M., N. Brener, and L. Kann (2010). Comparison of Paper-and-Pencil Versus Web Administration of the Youth Risk Behavior Survey (YRBS): Participation, Data Quality, and Perceived Privacy and Anonymity. *Computers in Human Behavior* 26(5): 1054–60

Denscombe, M. (2009). Item Non-response Rates: A Comparison of Online and Paper Questionnaires. *International Journal of Social Research Methodology* 12(4): 281–91

Dever, J. A., A. Rafferty, and R. Valliant (2008). Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias? *Survey Research Methods* 2(2)

Dey, E. L. (1997). Working with Low Survey Response Rates: The Efficacy of Weighting Adjustments. *Research in Higher Education* 38(2): 215–27

Faas, T. and H. Schoen (2009). Fallen Gewichte ins Gewicht? Eine Analyse am Beispiel dreier Umfragen zur Bundestagswahl 2002. *Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung.* Wiesbaden: VS Verlag für Sozialwissenschaften (145–57)

Franco, A., N. Malhotra, G. Simonovits, and L. J. Zigerell (2017). Developing Standards for Post-Hoc Weighting in Population-Based Survey Experiments. *Journal of Experimental Political Science* 4(2): 161–72

Fricker, R. D. and M. Schonlau (2002). Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods* 14(4): 347–67

Grewenig, E., P. Lergetporer, L. Simon, K. Werner, and L. Woessmann (2018). *Can online surveys represent the entire population?* CESifo working paper series No. 7222. CESifo Group Munich

Groves, R. M. and M. P. Couper (1998). *Nonresponse in Household Interview Surveys*. New York: Wiley

Groves, R. M. and L. Lyberg (2010). Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly* 74(5): 849–79

Hawkins, J. et al. (2013). Does gender influence participation? Predictors of participation in a community health worker diabetes management intervention with African American and Latino adults. *The Diabetes educator* 39(5): 647–54

Holbrook, A. L. and J. A. Krosnick (2010). Social Desirability Bias in Voter Turnout Reports. *Public Opinion Quarterly*(74): 37–67

Hoogendorn, A. and J. Daalmans (2009). Nonresponse in the Recruitment of an Internet Panel Based on Probability Sampling. *Survey Research Methods*(3): 59–72

Hoonakker, P. L. T. and P. Carayon (2009). Hoonakker, P.L.T. & Carayon, p. (2009). Questionnaire Survey Nonresponse: A Comparison of Postal Mail and Internet Surveys. International Jour-

nal of Human Computer Interaction, 25(5): 348-373. *International Journal of Human-Computer Interaction* 25: 348–73

Hugi, S. (2014). *Validierungsstudie Zu Den Verzerrungen von Selbstberichteten Politischen Partizipationsangaben in Der Schweizer Umfrageforschung*. Masterarbeit. Universität Bern

Hugi, S. (2014). *Validierungsstudie Zu Den Verzerrungen von Selbstberichteten Politischen Partizipationsangaben in Der Schweizer Umfrageforschung*. Masterarbeit. Universität Bern

Kalton, G. and I. Flores-Cervantes (2003). Weighting Methods. *Journal of Official Statistics*(19): 81–97

Keeter, S., C. Kennedy, M. Dimock, and J. B. P. Craighill (2006). Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey. *Public Opinion Quarterly*(70): 759--779

Keeter, S., C. Miller, A. Kohut, R. M. Groves, and S. Presser (2000). Consequences of Reducing Nonresponse in a National Telephone Survey. *The Public Opinion Quarterly* 64(2): 125–48

Klassen, R. D. and J. Jacobs (2001). Experimental Comparison of Web, Electronic and Mail Survey Technologies in Operations Management. *Journal of Operations Management* 19(6): 713–28

Kohut, A., S. Keeter, C. Doherty, M. Dimock, and L. Christian (2012). *Assessing the Representativeness of Public Opinion Surveys*. Pew Research Center for the People and the Press

Kreuter, F., S. Presser, and R. Tourangeau (2009). Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity. *Public Opinion Quarterly* 72(5): 847–65

Kruskal, W. and F. Mosteller (1979). Representative Sampling I: Non-Scientific Literature. *International Statistical Review*(47): 13–24

Kwak, N. and B. Radler (2002). A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics* 18(2): 257

Kwak, N. and B. Radler (2002). A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality. *Journal of Official Statistics* 18(2): 257

Leece, P. et al. (2004). Internet Versus Mailed Questionnaires: A Controlled Comparison (2). *Journal of Medical Internet Research* 6(4): e39

Little, R. J. A. and D. B. Rubin (2019). *Statistical Analysis with Missing Data. 3rd Ed.* New York: Wiley

Lundquist, P. and C.-E. Särndal (2013). Aspects of Responsive Design with Applications to the Swedish Living Conditions Survey. *Journal of Official Statistics* 29(4): 557–82

Manfreda, K. L., M. Bosnjak, J. Berzelak, I. Haas, and V. Vehovar (2008). Web Surveys Versus Other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research* 50(1): 79–104

McCabe, S. E. (2004). Comparison of Web and Mail Surveys in Collecting Illicit Drug Use Data: A Randomized Experiment. *Journal of Drug Education* 34(1): 61–72

Milic, T., M. McArdle, and U. Serdült (2016). *Attitudes of Swiss citizens towards the generalisation of e-voting*. No. 10. Studienberichte Des Zentrums für Demokratie Aarau. Aarau: Centre for Democracy Studies Aarau (ZDA)

Milic, T., U. Serdült, and S. Brüggemann (2018). *Studie zur kantonalen Volksabstimmung vom 23. September 2018*. No. 1. FOKUS Aargau. Aarau: Zentrum für Demokratie Aarau (ZDA)

Milic, T., U. Serdült, and S. Brüggemann (2019a). *Studie zur kantonalen Volksabstimmung vom 25. November 2018*. No. 2. FOKUS Aargau. Aarau, Switzerland: Zentrum für Demokratie Aarau (ZDA)

Milic, T., U. Serdült, and S. Brüggemann (2019c). *Studie zur Aargauer Nationalratswahl vom 20. Oktober 2019*. No. 5. FOKUS Aargau. Aarau: Zentrum für Demokratie (ZDA)

Milic, T., U. Serdült, and S. Brüggemann (2019b). *Studie zur Aargauer Ständeratswahl vom 20. Oktober 2019*. No. 4. FOKUS Aargau. Aarau: Zentrum für Demokratie (ZDA)

Norris, P. (2001). *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge: Cambridge University Press

R. M. Groves, J. F. F. Couper, M. P. Lepkowski, E. Singer, and R. Tourangeau (2004). *Survey Methodology*. Hoboken, NJ: John Wiley & Sons

Roberts, C., D. Joye, and M. E. Stähli (2016). Mixing Modes of Data Collection in Swiss Social Surveys: Methodological Report of the LIVES-FORS Mixed Mode Experiment. *LIVES Working Papers* 2016: 1–42

Rookey, B. D., S. Hanway, and D. A. Dillman (2008). Does a Probability-Based Household Panel Benefit from Assignment to Postal Response as an Alternative to Internet-Only? *Public Opinion Quarterly* 72(5): 962–84

Rosenbaum, P. R. and D. B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70(1): 41–55

Schnell, R. (1993). Die Homogenität Sozialer Kategorien Als Voraussetzung für „Repräsentativität" Und Gewichtungsverfahren. *Zeitschrift für Soziologie* 22(1): 16–32

Schnell, R. (1993). Die Homogenität Sozialer Kategorien Als Voraussetzung für „Repräsentativität" Und Gewichtungsverfahren. *Zeitschrift für Soziologie* 22(1): 16–32

Schouten, B., F. Cobben, and J. Bethlehem (2008). Indicators for the Representativeness of Survey Response. *Survey Methodology* 35

Sciarini, P. and A. C. Goldberg (2016). Turnout Bias in Postelection Surveys: Political Involvement, Survey Participation, and Vote Overreporting. *Journal of Survey Statistics and Methodology* 4(1): 110–37

Selb, P. and S. Munzert (2013). Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys. *Electoral Studies* 32(1): 186–96

Serdült, U., T. Milic, and S. Brüggemann (2019). *Studie zur Aargauer Regierungsratsersatzwahl vom 20. Oktober 2019*. No. 3. FOKUS Aargau. Aarau: Zentrum für Demokratie (ZDA)

Szolnoki, G. and D. Hoffmann (2013). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Wine Economics and Policy*(2): 57–66

Tourangeau, R., F. Conrad, and M. P. Couper (2013). *The Science of Web Surveys*. Oxford: Oxford University Press

Tourangeau, R., K. Rasinski, J. Jobe, T. Smith, and W. Pratt (1997). Sources of Error in a Survey of Sexual Behavior. *Journal of Official Statistics* 13(4): 341–65

Truell, A. D., J. E. Bartlett, and M. W. Alexander (2002). Response Rate, Speed, and Completeness: A Comparison of Internet-Based and Mail Surveys. *Behavior Research Methods, Instruments, & Computers* 34(1): 46–9

Wolfe, E. W., P. D. Converse, O. Airen, and N. Bodenhorn (2009). Unit and Item Nonresponses and Ancillary Information in Web- and Paper-Based Questionnaires Administered to School Counselors. *Measurement and Evaluation in Counseling and Development* 42(2): 92–103

Yeager, D. S. et al. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples. *Public Opinion Quarterly* 75(4): 709–47

Zhang, C., M. Callegaro, and M. Thomas (2008). More Than the Digital Divide? Investigating the Differences Between Internet and Non-Internet Users on Attitudes and Behaviors

Imprint

FOKUS Aargau is a research project of the Centre for Democracy Studies Aarau (ZDA). The FOKUS Aargau studies examine voting behaviour in cantonal referendums and elections. The project is financed by the Swisslos Fund of the Canton of Aargau. The data is collected by the survey institute *publitest* from Zurich, while the analyses are carried out by the ZDA. The final reports are available five to six weeks after the ballot date on the website reports.fokus.ag, further analyses can be found at publications.fokus.ag. The questionnaires underlying the studies are freely accessible on the same page.

Project responsibility

Uwe Serdült and Thomas Milic

Authors of the present study

Thomas Milic, Uwe Serdült und Salim Brüggemann, Zentrum für Demokratie Aarau (ZDA)

Citation suggestion

Thomas Milic, Uwe Serdült und Salim Brüggemann (2020): «What is the price of going fully online? A mixed-mode sampling experiment on the occasion of the Swiss general elections 2019». *FOKUS Aargau Special*. Aarau, Zentrum für Demokratie Aarau.

ISSN: 2624-7399
ISBN: 978-3-906918-09-9

Contact

Zentrum für Demokratie Aarau (ZDA)
FOKUS Aargau
Villa Blumenhalde
Küttigerstrasse 21
CH-5000 Aarau

info@fokus.ag

www.fokus.ag